Augmenting Photographs with Audio

David Frohlich, Guy Adams, Ella Tallyn[1]
Printing Imaging Technologies Laboratory
HP Laboratories Bristol
HPL-2000-105
9th August, 2000*

audio, photographs, audiophotos, tangible sound, augmented reality, interfaces

We report a user requirements study of several interfaces for the playback of sounds from photographs. The study showed that users liked listening to audiophotos when the sounds are played back from photographic prints, but as a compliment to playback on a PC. When handling prints the audio needs to be invoked manually from the print with a facility to pause the audio during playback. A handheld audioprint player was then designed to fulfill these needs, based on an embedded chip in the paper.

[1] Flow Interactive Ltd
* Internal Accession Date Only
Approved for External Publication

1. Introduction

The value of ambient sounds with photographs was established in an audiocamera field trial in which several families were encouraged to capture combinations of sound clips and photographs on the same device [1]. Ambient sounds-of-the-moment were the most common and attractive type of sound captured, serving to give mood and life to the photo and trigger a richer remembering of the event. The issue of how users want to playback audiophoto material was not addressed in our previous paper, although this is critical for the design of technology to enable a mass-market practice of audiophotography. In this paper, we describe our attempts to understand the user needs for audiophoto playback, and a resulting invention to satisfy those needs.

An obvious way to playback audiophoto material captured on a digital camera is to display the photos on a PC, TV or portable viewing screen (such as the camera itself) and have the associated sounds play automatically or under manual control. Indeed this is the method currently used on existing digital cameras with sound capture facilities such as the Ricoh RDC-4300 , the Kodak DC260 and the Sony DSC-F55. However, we were intrigued by the possibility of an alternative 'augmented reality' approach in which sounds might be played back from printed photographs or 'audioprints'. In this, we were inspired by a number of augmented reality interfaces in which sounds are evoked from physical objects, such as Durrell Bishop's **Marble Answering Machine** [2] and Lisa Stifelman's **Audio Notebook** [3]. Furthermore we were aware of a number of commercial products which simulate audioprints, such as **recordable photo frames** (e.g. Magic Moments) and **greeting cards** (Hallmark), an audiocard reader called **Language Master** (Bell & Howell), voice message holders such as **FamilyVoice** (Cannon) and the **ScanTalk** pen (Olympus) which reads printed sound labels.

In order to understand the relative user value of conventional and augmented reality approaches to audiophoto review, we used a mixture of existing technology props and semi-functional prototypes as 'conversation pieces' for eliciting creative user feedback and requirements.

2. Methods

At home visits in the original audiocamera trial we asked four families to review packs of their own loose printed photos whilst listening to a serial playback of the corresponding audio from a **cassette tape**. This meant that they had some initial experience of audio playback from a stack of prints. At the same home visit we asked families to select their favorite audiophotos for display in a family album. Pairs of families were later invited into a usability suite in HP Labs Bristol to review this material, presented in a variety of ways. Three key methods were an audio CD, a PC audiophoto album and a handheld audioscanner mock-up.

The **audio CD** method was a variant of the cassette tape solution. Printed photographs had audio track numbers written on the back. These audio tracks could be randomly accessed from a standalone hi fi unit nearby (actually a minidisc player). The **PC album** presented up to three 6" X 4" images per screen page. Users had to click on each photo to playback the corresponding sound clip, and could use forward and backward buttons to move between pages at will. The **audioscanner** simulated a portable sheet feed scanner designed to read encoded audio information from the back of the print, for playback in the hand. It comprised a box measuring 5.5" X 3" X 2" with a 4" slot in the front, a loudspeaker in the top and a standard audio cable and stereo jack protruding from the side (see Figure 1). This was used in conjunction with a minidisc player to playback the audio clip corresponding to printed photo sample pushed into the slot.

Figure 1. The audioscanner mock-up

3. Results

By observing audio playback from the **cassette tape** in the home visits, we found that families automatically modified the way they handled and discussed the photos in order to accommodate the audio (see Figure 2). They switched spontaneously from passing individual photos along the family group, to a process of one person turning over photos in the pack. This led to a more intimate experience, but one which removed individual choice of how long to look at each photo. In fact, the photo holders tried hard to turn over the photos in pace with the sounds, to preserve the correspondence between image and audio. This sometimes caused the families to dwell on some audiophotos for longer than they wanted to, and to speed through photos without sounds too quickly. As one participant said:
Phillip: There is no time to linger over the photos ..at the tape pace (Visit 2)

Curiously, the same effects were observed later when families used the **audio CD** method. Despite the ability to pause the audio or randomly access audio clips out of sequence on the hi fi, photo holders tended to turn over the pack systematically and let the audio clips play through. They simply weren't able to manually control both the photos and the hi fi unit simultaneously:
Mervyn: It was too fiddly (Group 1)

Reviewing audiophotos on the **PC** led to a more orderly process in which one PC user clicked on individual photos in sequence to play the corresponding sounds. The fact that a manual action was required to invoke individual sound clips meant that there was more opportunity to talk about the photos in between the clips, and to control the timing and duration of clips. Furthermore this control was shared by the rest of the review group who clustered around the screen and shouted out instructions to pause or replay the sounds, or skip onto the next page of the album. In some cases observers actually grabbed the mouse to perform the control actions themselves. This led to a general feeling that the PC album looked 'professional' but was easy to use:
Chris: It looked impressive and you can trawl through them easily (Group 2)
Liz: It's so easy to use, there's an option to have sound or you can just flick through (Group 1)

(a)

(b)

Figure 2. The effect of playing a continuous audio soundtrack on photo handling: (a) photo passing without audio, (b) photo turning with audio

Although families couldn't handle and interact with the **audioscanner** mock-up as they could with the prints and PC album, they immediately saw it as providing PC-like control for prints:
Christine: You've got the choice of showing just the photos (Group 2)
Thomas: You can go to the back one without pressing the button all the time (Group1)
In addition, the audioscanner was valued for its portability, simplicity and compatibility with existing tangible prints:
Christine: Its small and mobile and the photos and sounds are together (Group 2)
Sue: I want to show them casually where I am at the time. You could sit out in the garden (Group 1)
Will: Anyone could use it. I mean if you were taking it to granny and grandpa (Group 1)
Debbie: There is something nice about having that tangible print in your hand (Group 2)

To consolidate these discussions of alternative review methods we asked individual family members to arrange the playback methods in order of preference. Across the group we found that most people ranked the PC album and the audioscanner as their top two choices, with roughly equal numbers putting the PC first as putting the scanner first. In conversation later, this majority said they wanted access to **both** methods of playback. The print/tape and print/CD combinations were discounted as impractical.

4. Implications for design

The findings show that playback of sounds from printed photos is seen as an attractive extension to PC display, as long as the pace of the sounds can be carefully controlled in conversation. In this respect our audioscanner mock-up turned out to be a good design functionally, because a manual action on the photo triggers only the sound for that photo. However, a subsequent technical investigation suggested that encoding the sound data in printed form did not provide a scalable or cheap enough solution. For example, the capacity of sound that could be stored with the photo is limited to what can be printed on the rear surface area of a 6" X 4" photo, while the cost of the scanner is likely to be unacceptably high.

Instead we chose to follow-up these findings by building a handheld audioprint player which utilises sound stored on a miniature audio chip embedded in the photo (see Figure 3). The technology chosen has several key values. The storage chip is barely thicker than the print itself and with further effort would be the same or thinner. The player is simple and would be cheap to produce in modest volume. It has simple play, pause, rewind and volume controls meaning operation is intuitive. It is also capable of recording the sound clips via a simple audio jack from any source, or can capture annotation via a built in microphone. Future research is directed at testing this kind of solution against others in a field trial situation.

Figure 3. The audioprint player

Finally, the fact that not all types of tangible sound interfaces were rated equally in the study, and that the PC interface was also rated highly sends a strong message to designers of augmented reality interfaces. In arguing their case they should be careful to contrast the perception and performance of their designs with those of more conventional alternatives, and to follow the kind of user-centered design process necessary to improve the usability of any interface or interaction technique. This will involve taking a stronger user focus than has typically been adopted in the field, with greater attention to the application domain and its particular requirements, and to the details of implementation and their effects on behavior.

Acknowledgements

Thanks to Phil Neaves and Simon Poole for making up the audioscanner device and Graham Allen Associates for production of the audioprint player.

References

1. Frohlich, DM & Tallyn E. Audiophotography: Practice and prospects. Extended Abstracts of CHI ’99 Conference on Human Factors in Computing Systems, 1999; 296-297.
2. Crampton-Smith G. The hand that rocks the cradle. ID Magazine, May/June 1995; 60-65.
3. Stifelman L. Augmenting real-world objects: A paper-based audio notebook. Companion proceedings of CHI ’96 Conference on Human Factors in Computing Systems, 1996; 199-200.